**Request for Information (RFI) SN08-52:** Computer Learning of Linguistic Representations of the Physical World

The Defense Advanced Research Projects Agency (DARPA) Information Processing Techniques Office (IPTO) is requesting information on areas of research relating to the enabling of computer learning of linguistic representations of the physical world.

Although human language may be described as inherently symbolic and arbitrary, it is an effective means of communication. Humans are extremely adept at associating linguistic symbols with the physical entities and abstract phenomena that the symbols represent. Despite the fact that, in some aspects, computer processing power already exceeds that of a human brain, current computer technology does not come close to matching the symbol association ability of even a very young human. In general, previous attempts at enabling computers to automatically acquire language understanding have been limited to attempts at creating computer analysis of syntax, word functions, and predefined relations. There have been some manual efforts to construct databases to encode meanings, but doing so required a great deal of labor. Furthermore, when some information is missing from a large database, it is almost impossible to add this information automatically. Manual computer systems also lack the ability to learn automatically and dynamically, as is required by real-world situations.

DARPA/IPTO is soliciting ideas and methodologies for both breaking this barrier of computer language understanding, and overcoming computers' inability to associate linguistic constructs with real world objects, actions, and concepts. Specifically, DARPA/IPTO is interested in descriptions of current or proposed research on creating technology enabling computers to acquire the ability to encode linguistic representations of physical entities and actions at the level of a 2-3 year old human. Such encoding ability will allow more flexible and powerful knowledge representations about the physical world; "self improvement," including automatic best-effort classification of new input; development of powerful mixed-initiative dialog ability; learning of new linguistic constructs; and the creation of generalized knowledge structures (e.g. ontologies or similar hierarchical taxonomic representations). Further, the ability to create linguistic representations of the physical world will allow the encoding and storage of properties of the entities that words represent in order to enable acting, reasoning, and planning through computer technology.

As this topic encompasses multiple scientific disciplines, DARPA/IPTO welcomes white paper contributions not just from the fields of artificial intelligence and linguistics, but also from the fields of: vision (including perception of real-world 2D and 3D scenes, as well as schematized or "cartoon" input), developmental psychology, neuroscience and childhood education. Of particular interest are educational approaches (including targeted curricula), teaching materials and pedagogical strategies for teaching a computer about linguistic representations of the world. Areas for consideration include, but are not limited to: (1) novel forms of internal knowledge representation; (2) innovative ideas with respect to the acquisition of linguistic representations, including both supervised and unsupervised methods; (3) mappings and other techniques of associating concepts, words, facts, and situations contained in external natural language media

with internal computer representations, and; (4) evaluation methodologies for computer language learning technology.

**WORKSHOP**

A DARPA-sponsored workshop has been planned for 2-3 December 2008 in Northern Virginia, for the purpose of reviewing on-going research in machine learning and artificial intelligence. Information presented at this workshop will contribute to the formulation of possible future areas of DARPA research with the objective of creating prototype systems that can enable computers to acquire the ability to encode linguistic representations of physical entities and actions at the level of a 2-3 year old human.

Space for the workshop is limited and attendance will be by invitation only. **Invitations will be based on white papers submitted, per the instructions below, no later than 1200 noon (ET), 23 October 2008.** Accepted authors will be notified via email by 28 October 2008, and at that time will be given further details on the workshop (times, location, etc.). In addition to being invited to the workshop, some authors may be invited to provide a 15-25 minute (with five minutes for questions), unclassified, non-proprietary presentation on related work and on-going research activities associated with machine learning and intelligence. If there are more interesting presentations submitted than time allows, a poster session may be added. All attendees will be encouraged to participate in general discussions on the papers and make recommendations for future research in the area.

**WHITE PAPER SUBMISSION INSTRUCTIONS AND FORMAT**

DARPA will employ an electronic upload process for responses to this RFI. To respond to the RFI, interested parties must complete an online cover sheet for each white paper response, which will include the information outlined below. The coversheet submission site is https://www.csc-ballston.com/rfi/rfiindex.asp?RFIid=08-52. Upon completion of the online cover sheet, a confirmation screen will appear, along with instructions on uploading the white paper.

**Since candidate authors may encounter heavy traffic on the web server, they SHOULD NOT wait until the day submissions are due to fill out a coversheet and submit the white paper!**

DARPA will acknowledge receipt of submissions via email within 3 business days of the deadline.

White papers should briefly summarize approaches, and not exceed 5 pages, including figures. Format specifications include 12 point font, single-spaced, single-sided, 8.5 by 11 inches paper, with 1-inch margins in either Microsoft Word or Adobe PDF format and zipped with either Winzip or PKZip.

1. Cover Page (1 page)
    a. Title of paper
    b. Organization
    c. Respondent's technical and administrative points of contact (names, addresses, phones and fax numbers, and email addresses)

d. Indication of willingness to attend the Workshop
e. Indication of willingness to present at the Workshop

2. Technical Ideas (up to 4 pages)
    a. Executive summary
    b. What the seed knowledge for such a project will be
    c. What the basic technologies would be needed to enable a computer to perceive information from the physical world and represent it linguistically. Included should be a description of:
        i. What form of input the computer must receive
            - Text
            - Still pictures (2 or 3D)
            - Motion pictures (2 or 3D)
            - Binocular vision
            - Robotic tactile sense
        ii. What level of human interaction with the computer would be necessary
        iii. What types of underlying representations would accomplish this ability
    d. What capabilities the computer would attain and how it would demonstrate these capabilities. For example:
            - Answering questions
            - Carrying on dialogues
            - Interacting through actions
    e. How progress would be measured

Respondents are encouraged to be as succinct as possible while at the same time providing actionable insight.

**ELIGIBILITY**

DARPA invites participation from all those engaged in related research activities and appreciates responses from all capable and qualified sources including, but not limited to, universities, university-affiliated research centers, federally-funded research centers, private or public companies and Government research laboratories.

**DISCLAIMERS AND IMPORTANT NOTES**

This is an RFI issued solely for information and new program planning purposes; the RFI and workshop do not constitute a formal solicitation for proposals. In accordance with FAR 15.201(e), responses to this notice are not offers and cannot be accepted by the Government to form a binding contract. Submission of a white paper, and/or attendance at the workshop, is voluntary and is not required to propose to subsequent Broad Agency Announcements (if any) or research solicitations (if any) on this topic. DARPA will not provide reimbursement for costs incurred in responding to this RFI or participating in the RFI workshop. NO PROPRIETARY OR CLASSIFIED INFORMATION SHALL BE INCLUDED IN THE RFI RESPONSE. Respondents are advised that DARPA is under no obligation to acknowledge receipt of the information received, or provide feedback to respondents with respect to any information submitted under this RFI.

Submissions may be reviewed by: the Government (DARPA and partners); Federally Funded R&D Centers (such as MIT Lincoln Laboratory); and Systems Engineering and Technical Assistance (SETA) contractors (such as Schafer Corporation, Science and Technology Associates, CACI International, and System Analysis, Inc.).

**POINT OF CONTACT**

Dr. Joseph P. Olive, IPTO Program Manager, DARPA, Email SN08-52@darpa.mil. ANY INQUIRIES ON THIS RFI AND/OR WORKSHOP MUST BE SUBMITTED TO SN08-52@darpa.mil. NO TELEPHONE INQUIRIES WILL BE ACCEPTED.